Study of Usability and Effectiveness in Shape and Shape-Color Geometric Shape Pattern-Based CAPTCHA

Thawatwong Lawan

Abstract —**This study investigates usability and effectiveness in shape and shape-color Geometric Shape Pattern-base CAPTCHA. The purpose of this study is to compare the amount of samples, sex and human generations on usability and effectiveness of the developed Geometric Shape Pattern-base CAPTCHA. There are 400 participants who approaching this study on the internet. The findings suggest that usability shown in percentage of success, both shape and shape-color Pattern-base CAPTCHA show highest percentage of success at 95.83% and 95.64% respectively. Moreover, it is demonstrated that the amount of time spent for finding the correct answer relates to the amount of sample of series. In addition, differences for sex reveal that both female and male can solve two types of CAPTCHA similarly. However, differences for human generations suggest that Baby Boomer I spends 4.63 to 7.38 seconds less to solve the problem of Pattern-base CAPTCHA. Moreover, in color-shape pattern, they spend 3.88 to 6.72 seconds less.**

Index Terms—**CAPTCHA, geometric shape, shape pattern.**

I. INTRODUCTION

The main purpose of CAPTCHA is firstly, to automate Public Turning Test in order to separate computers and human being apart such as their names and roles. This can be done by using a challenge-response test. It is used in computing to determine whether the user is human or not. [1] Second main purpose of CAPTCHA is it is the first protector that guards the front gate of the computer's screen. It will decide who is allowed to pass the gate.

These purposes are most important roles of CAPTCHA. That is because for example, if the guard of the house is frigid, unfriendly, cruel and cannot be able to communicate effectively; the guest will get angry or would like to stop contacting to the house-owner. Thus, the house-owner cannot access to what they want such as news, information, and he may not be able to contact to the others. On the other hand, if the guard of your house is ineffective, credulous and unable to screen everyone who want to access the house's gate; the house will be full of unscreened people. As the result of that, the house-owner will be put in danger. Furthermore, CAPTCHA is highly effective for being a good protector for your computer. It can be accessed easily for real human. It is also can be used by multinational

people. Although CAPTCHA is easy to use, it can be easy only for real human. CAPTCHA is hard to access for computer bots duplicating or disguising as a real human to access into computer's important database. Recently, programmers develop many kinds of CAPTCHA such as Text-based CAPTCHA, image-based CAPTCHA etc. [2], [3]. Text-based CAPTCHA was the primary proposed scheme. They use warping and distraction techniques for preventing bots. It seems difficult for automatic bots to pass but they also appear the same problem to human [4], [5]. However, this type of CAPTCHA become popular and used until now. Even though it's warping was still a significance drawback. The sample of text-based CAPTCHA is shown in Fig. 1.
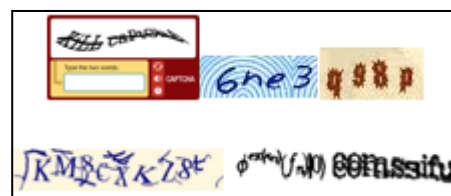

Fig. 1. Text-based CAPTCHA.

In addition, some programmers develop Audio-base and Image-base CAPTCHA. These CAPTCHAs are difficult for generating and evaluation [6]. Moreover, image-base CAPTCHA is need to store in big databases and need more time in generating phase [7]-[9]. The puzzle and Multi-Modal CAPTCHA [10], [11] claimed that they overcome language limitations. But we still found that it may use more time in implementation phase and some puzzles harder for the user to solve. Furthermore, all kinds of traditional CAPTCHAs have some problems about communication to user including problem about language or knowledge of the approach.

Pattern is a series of number or figure that arranged by following a rule or rules [12], [13] and it very simple in Mathematics. Everybody used to solve patterns such as color pattern, shape pattern and number pattern. Patterns are everywhere around us. Therefore, finding and understanding patterns can give us more experiences in solving them. Patterns can help us learn to predict the future, discover new things and understand this world better [14]. In consequence of these reasons, the researcher develops geometric pattern-base CAPTCHA and implement in the approach. The result of previous study showed that the Geometric shape Pattern-base CAPTCHA can fulfill the gap that is the problem of all CAPTCHA, which is the problem of communication for multinational people. That is because Geometric shape Pattern-base CAPTCHA does not require any particular language to communicate. Geometric shape Pattern-base

CAPTCHA can communicate to everyone by themselves. The correct answer will come out from their observation, analysis and conclusion of pattern relating to pattern series that people acquire. This is the strength of Geometric shape Pattern-base CAPTCHA to fight against computer bots because they do not have thinking process similar to human. However, in the first presented Geometric shape Pattern-base CAPTCHA, the researcher uses only one factor to build Geometric shape Pattern-base CAPTCHA. In this study, the researcher adds more factors in Geometric shape Pattern-base CAPTCHA to compare the result of usability and effectiveness with the previous factor Geometric shape Pattern-base CAPTCHA.

II. METHODOLOGY

A. Population and Samples

In this study, methodology refers to the approaching used in the environment of open internet system for everyone who would like to participate. The amount of samples refer to the Electronic Transactions Development Agency (Public Organization), or ETDA survey result of amount of Thailand Internet User Profile 2016 were 38,015,725. Then, the sample's size is calculated by TARO YAMANE [15] formula as shown in equation 1.

$$n = \frac{N}{(1+Ne^2)} \tag{1}$$

where;

n = Sample size
N = Population size
e = expected error size = 5%
So Sample size = $38,015,725/1+38,015,725(0.05)^2$
$= 400$

The methodology uses JAVA script and PHP on internet platform for development of Geometric shape Pattern-base CAPTCHA. There are two involved phases in the phase of the Geometric shape Pattern-base CAPTCHA webpage: general data and Geometric shape Pattern-base CAPTCHA. In the phase of Geometric shape Pattern-base CAPTCHA, the approach will be analyzed and find the answer of CAPTCHA. Then, they will evaluate themselves. There is no language to explain the relationship of pattern series.

B. Implementations

The Geometric shape Pattern-base CAPTCHA in this study are presented in research website. There are 6 Geometric shape Pattern-base CAPTCHA showed in Fig. 2.
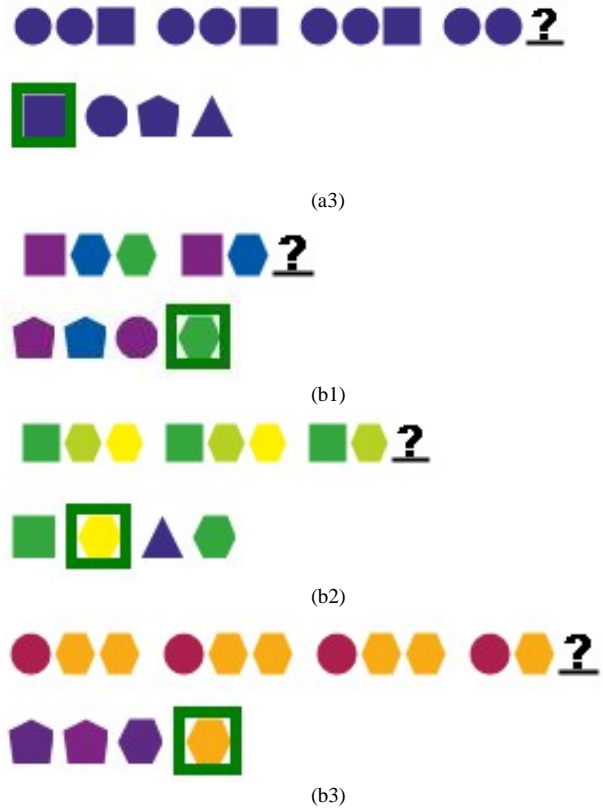

(a1)

(a2)

Fig. 2. Characteristic of Geometric-shape Pattern-based CAPTCHA. (a1)-(a3) as a 1-sample,2-sample and 3-sample of Shape Pattern-based (b1)-(b3) as a 1-sample,2-sample and 3-sample of Shape-Color Pattern-based CAPTCHA.

One factor CAPTCHA is Geometric shape Pattern-base CAPTCHA using only shape factor in this CAPTCHA. At the same time, two factor CAPTCHA is Geometric shape Pattern-base CAPTCHA using shape and color factors. There are three levels in both One factor CAPTCHA and Two factor CAPTCHA: 1) Level one has one sample of series. 2) Level two has two sample of series and 3) Level three has three sample of series.

The approach will observe and try to comprehend the relation of pattern series. Then, it will choose the correct answer. There is no limit of time to find answers for all of the questions. The researcher will collect data about the amount of time to answer the questions, number of all answers, number of first correct answers, and numbers of wrong answers. And then calculate the Geometric shape Pattern-base CAPTCHA performance by percent of success. In the Geometric shape Pattern-base CAPTCHA usability show in time to pass the CAPTCHA compare in sex and generation age in second unit. All of data will collected for Statistical Analysis.

III. RESULT AND DISCUSSION

Data analysis in this study are amount, percentage, means and standard deviation and statistic for analysis using program for statistic analyze SPSS for Windows Version 10.0.

The result of the study of usability and effectiveness in pattern and pattern-color Geometric Pattern series is shown in 2 tables and 2 figure below:
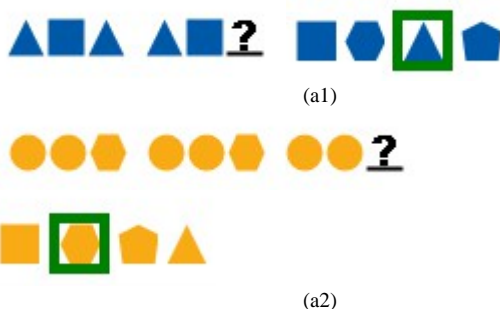
TABLE I: NUMBER OF FIRST CORRECT, NUMBER OF TOTAL ANSWER AND PERCENTAGE OF SUCCESS IN DIFFERENCES AMOUNT OF PATTERN AND COLOR-PATTERN SERIES CAPTCHA'S SAMPLE

Type of CAPTCHA	The Number of			% of Success
	Samples	First Correct	Total Answer	
Pattern series	1	366	436	83.94
	2	391	414	94.40
	3	394	411	95.83
Color-pattern series	1	381	464	82.65
	2	391	429	91.14
	3	395	413	95.64

From the table, the result shows in pattern series that the amount of samples effect to percentage of success. If there are more samples in Geometric Pattern series CAPTCHA, percentage of success will be high. Moreover, the same result applies in shape-color pattern series too. This result suggests that Geometric shape Pattern-base series is the principle mathematic about pattern and series that everyone can solve by observing and analyzing how each sample shows relation and conclusion for the answer. The ability to find the correct answer depends on whether series has been seen before. Therefore, if there are more samples, the approach will find the correct answer faster.

This result supports that Geometric shape Pattern-base CAPTCHA does not have to communicate to the approach by using any language. Furthermore, in comparison of the type of Geometric shape Pattern-base CAPTCHA among shape series to shape-color pattern, the result shows percentage of success at 95.83% in shape pattern and 95.64% in shape-color pattern. It can be shown that the approach can understand and succeed to find the correct answer.

TABLE II: X-BAR ±SD. OF TIME TO FIND CORRECT ANSWER IN DIFFERENCE AMOUNT OF SAMPLE OF SHAPE AND SHAPE-COLOR PATTERN-BASE CAPTCHA

CAPTCHA type	No. of sample	$\bar{X} \pm$ SD. (sec)
Pattern series	1 sample	10.64 ±8.49
	2 sample	9.09 ± 8.19
	3 sample	7.98 ± 6.90
Color-pattern series	1 sample	6.16 ± 14.64
	2 sample	6.23 ± 9.29
	3 sample	6.27 ± 11.08

From the above table, it is shown that the approach spends less time to pass CAPTCHA in both Geometric shape Pattern-base CAPTCHA shapes pattern and shape-color pattern. The duration spent on finding the correct answer from shapes pattern and shape-color pattern are 7.98-10.64 seconds and 6.27-6.16 seconds respectively. In addition, the result also shows the same trend to percentage of success. That is more amount of samples will make the approach spends less time than less amount of samples. After all, the result indicates the strength of Geometric shape Pattern-base CAPTCHA due to the duration that the approach spends for finding the correct answer. This result indicates that more sample patterns will make the approach gets the hints easily. In queries about the attributes of things, the approach will apply reasons to the answer "What's next?" with a description not a number. Moreover, human will develop "function sense" without simply asking for the next
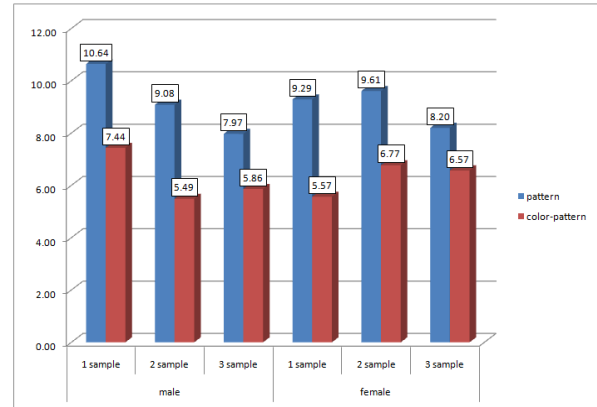
shape in the pattern [12], [14].

Fig. 3. Shows the duration to find correct answer in differences sex and amount of samples of shape and shape-color pattern-base CAPTCHA.

Fig. 3 shows the amount of time to find the correct answer. Differences for sex, the finding suggests that female spends time to find the correct answer at 8.20-9.29 seconds in shape pattern and 5.57-6.77 seconds in shape-color pattern. At the same time, male spends 5.49-7.44 seconds and 7.97-10.64 seconds in shape pattern and shape-color pattern respectively. Furthermore, differences of the amount of samples, the finding indicates the same result in both shape and shape-color pattern. If there are more amount of samples, the approach will use less time to find the correct answer. The observation point of this figure is that when the approach uses only one factor, it spends more time than using two factors: color and pattern. This result strongly supports the second reason. Sex is not barrier of human's function sense. Two more sample patterns effect on both sex to get hints that can solve the answer of patterns. That is because patterns pervade every parts of human's lives. In addition, through visual and written forms of expressions, all approach will connect language and mathematics to develop skills for thinking clearly, strategically, critically, and creatively [12].

In Fig. 4, human generations are divided into four generations: Generation Y, Generation X, Generation Jones, and Baby Boomers I. That is because each generation has faced same environment and experience about education system and technology concept, which may take some effects to internet or website used in this study.
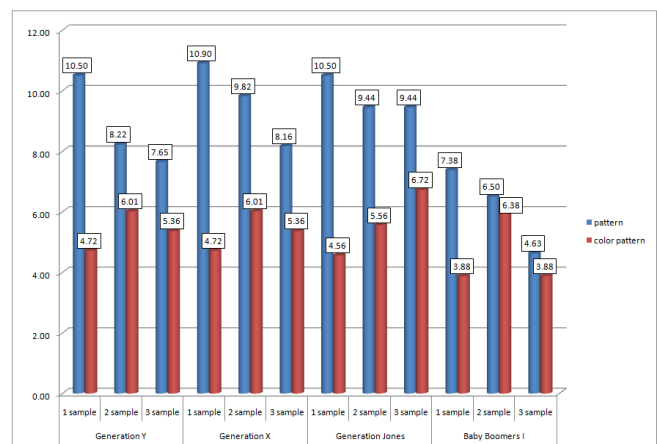

Fig. 4. Shows the amount of time to find the correct answer in differences generation ages and amount of samples of shape and shape-color pattern-base CAPTCHA.

The result indicates that Baby Boomers I spends less time to find the correct answer. They spend 4.63-7.38 seconds in shape pattern-base CAPTCH and 3.88-6.72 seconds in shape-color pattern. On the other hand, Generation Jones spends longest time to find the correct. In shape pattern-base CAPTCH, they spend 9.44-10.50 seconds and in color-shape pattern, they spend 4.56-6.72 seconds. This finding confirms that both generations and ages have no effect on time to find the correct answer. It is shown that differences of generation and ages can solve the correct answer of CAPTCHA. However, in older age, giving two factors of Geometric Shape Pattern-base CAPTCHA will make them get more hints. This finding is the interesting point for further study.

IV. CONCLUSION

Everybody learns and sees patterns around them from childhood to adulthood. Therefore, all of us develop function sense and we use this sense in more aspect in our lives. That is making human differentiates from a computer. Human learns to see, observe, and conclude for finding correct answer to choose everything in their lives. The Geometric shape Pattern-base CAPTCHA is one kind of CAPTCHA that tries to use natural sense of human, which is more effective than a computer. That is because it can demonstrate simple thinking process that can prove whether the user is a real human [12], [14]. This strength of Geometric shape Pattern-base CAPTCHA can be new highly efficient and alternative type of CAPTCHA for everyone. In addition, as this CAPTCHA do not require any language to communicate for finding the correct answer, they suggests high percentage of success; and are more effective to multinational people in every sex and human generations. The series sample will communicate to people. Moreover, this CAPTCHA do not require more amount of pattern samples. This study demonstrates that only two or three samples are enough for people to understand and get a hint of series' relationship, which can lead to the correct answer. However, we will test the automated bots ability to break down the challenges in the real world and we will try to vary more factors evaluation, such as shape's color, position of shape in the series, and the combination of both or we may be use the animated shapes as element in the series of pattern in our future work.

ACKNOWLEDGMENT

REFERENCES

[1] The re CAPTCHA Project–Carnegie Mellon University CyLab. [Online]. Available: www.cylab.cmu.edu

[2] S. Ved and P. Preet, "Survey of different Types of CAPTCHA," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, pp. 2242-2245, 2014.

[3] W. K. Abdullah Hasan, "A survey of current research on captcha," *International Journal of Computer Science & Engineering Survey (IJCSES),* vol. 7, no. 3, June 2016.

[4] G. Moy, N. Jones, C. Harkless, and R. Potter, "Distortion estimation techniques in solving visual captchas," in *Proc. the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* USA: IEEE Computer Society, 2004, pp. 23–28.

[5] C. Pope and K. Kaur, "Is it human or computer? Defending e-commerce with captchas," *IT Professional,* vol. 7, no. 2, pp. 43–49, Mar. 2005.

[6] R. Gossweiler, M. Kamvar, and S. Baluja, "What's up captcha?: A captcha based on image orientation," in *Proc. of the 18th International Conference on World Wide Web, Www'09,* ACM, 2009, p. 841.

[7] J. Elson, J. R. Douceur, J. Howell, and J. Saul, "Asirra: A captcha that exploits interest-aligned manual image categorization," in *Proc. ACM CCS,* 2007, pp. 366–374.

[8] J. Yan and A. S. E. Ahmad, "Usability of captchas or usability issues in captcha design," *International Conference Proceeding Series,* ACM, pp.44–52, 2008.

[9] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "Recaptcha: Human-based character recognition via web security measures," *Science*, vol. 321, no. 5895, pp. 1465–1468, 2008.

[10] F. Ali Bin Hamid Ali and F. Bt. Karim, "Development of CAPTCHA system based on puzzle," *I4CT*, pp. 426-428, 2014.

[11] A. S. Almazyad, Y. Ahmad, and S. A. Kouchay, "Multi-Modal CAPTCHA: A User Verification Scheme," in *Proc. 2011 International Conference on Information Science and Applications (ICISA),* Jeju Island, pp. 1-72011.

[12] The Annenberg Foundation. (2017). Pattern in mathmatics. [Online]. Available: https://www.learner.org/teacherslab/math/patterns/

[13] H. Komatsu and Y. Ideura, "Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey," *J Neurophysiol*, vol. 70, no. 2, pp. 677-94, 1993.

[14] Pattern. [Online]. Available: http://www.mathsisfun.com/algebra/patterns.html

[15] Thailand Internet user survey 2016. (Feb. 2017). Electronic Transactions Development Agency (Public Organization). [Online]. Available: https://www.quora.com/What-is-Yamane-sample-calculation

Thawatwong Lawan was born in 1968. He works as a lecturer at the Department of Computer Science, Faculty of Informatics Mahasarakham University, Thailand. He received bachelor's degree in computer science and master's degree in computer engineering. He taught computer networking, data communication, network programming, legal and ethical issues in computer science. He interested in wireless network, network security, quality of services, and apply of computer network.

Electronic System & Measurement and Control Technology